Speech Studio Users Guide

Laryngograph ®

1 Foundry Mews, London, NW1 2PE
Tel: +44 (0)207 387 7793 Fax: +44 (0)207 383 2039

lx@laryngograph.com www.laryngograph.com

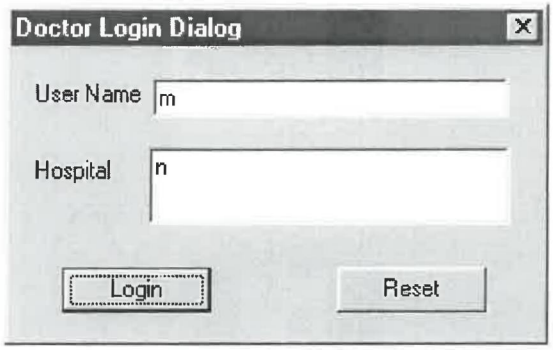
Contents

Getting Started	1
The Main Display Screen	2
Recording Data	3
Zoomed Wave Mode. Saving Data.	4
Opening a Saved File	5
Displaying Recorded Data. Zooming and Viewing.	6
Using Scale to Fit	8
Listening to the Recorded Data. Saving a Section of Data.	9
Selecting Displays.	10
Voice Pattern Display and Waveform Control Bar.	11
Target Modes	12
Jitter and Shimmer Analysis on Sustained Vowels	13
The Use of QA in the Analysis of Connected Speech	14
The Use of Set	15
Appendix	18

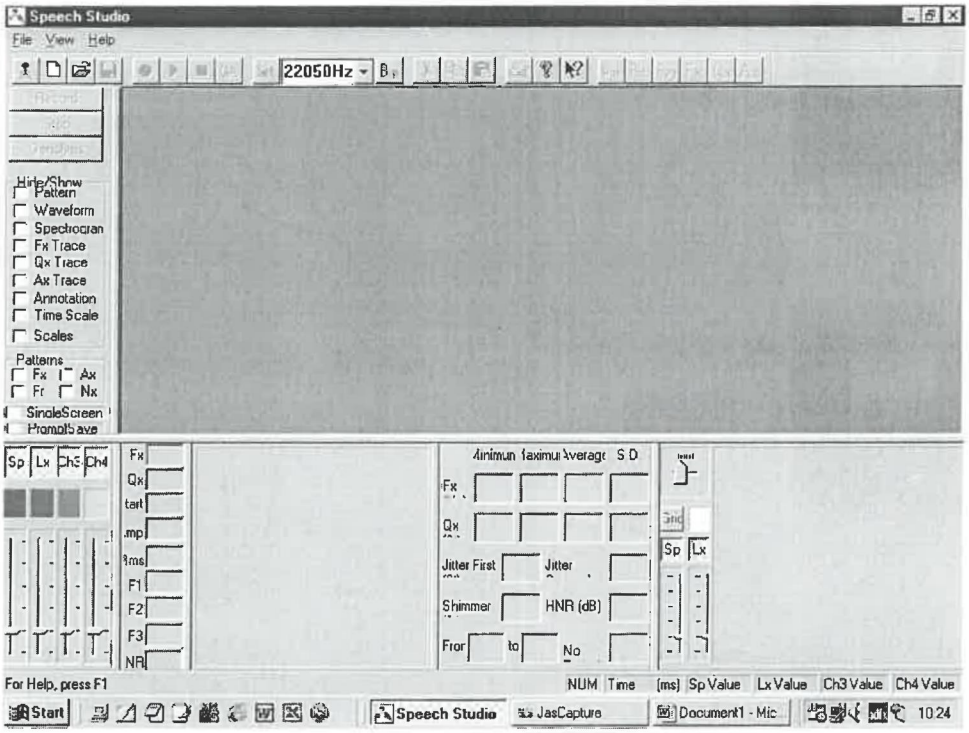
Getting Started

When Speech Studio is installed on your computer it will have its own icon on your screen. Double click on the **spead** (**s**peech **p**attern **e**lement **a**nalysis and **d**isplay) icon. The first time that you use Speech Studio you will get a new screen display that contains a composite view of some of the interactive display windows available.

Each time Speech Studio is used in a clinical session, a panel is displayed superimposed on the Main Display Screen – as on the right. The individual user can then enter their name and/or other identifying information before using the system with patients. This information will appear on any printed reports.



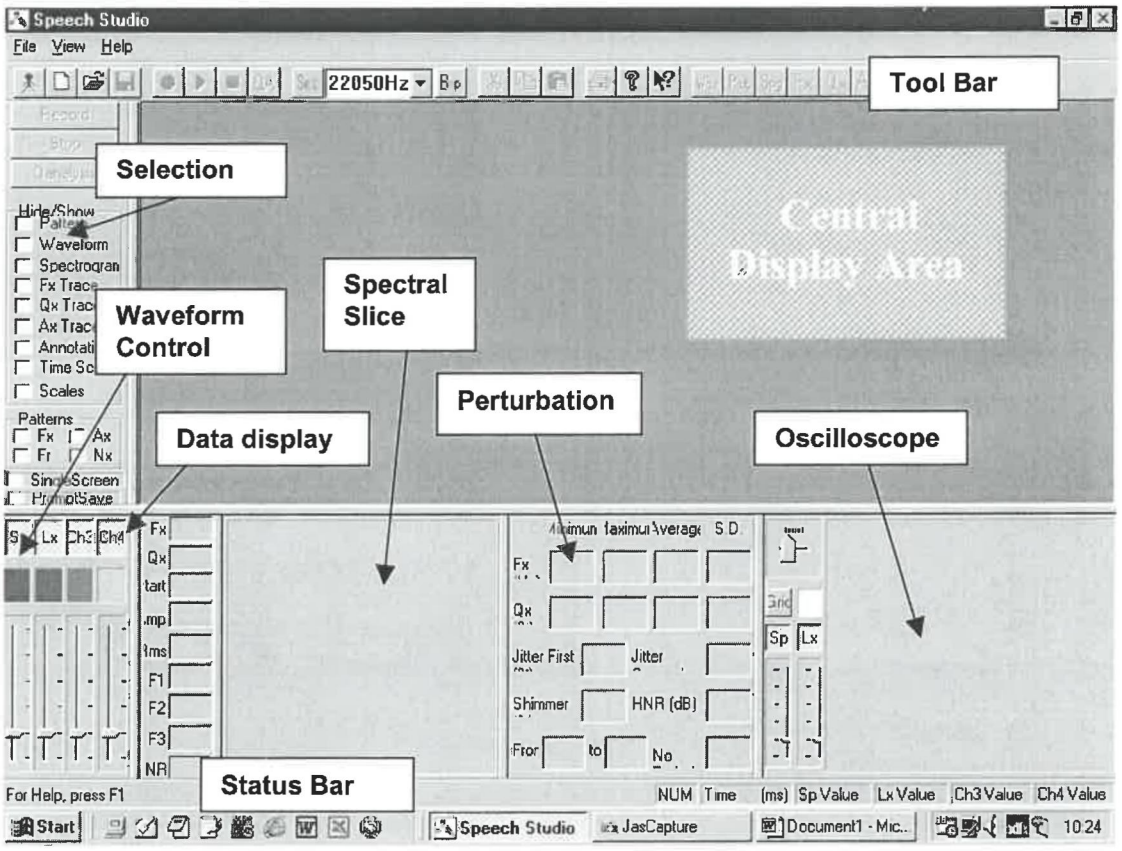
The Main Display Screen



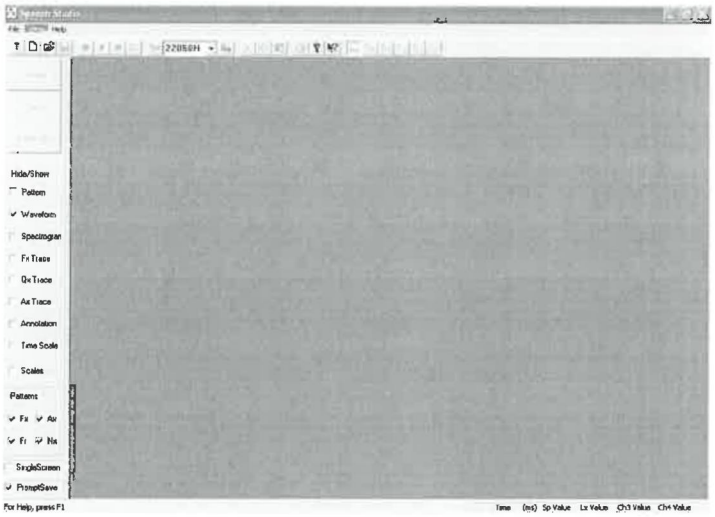
The screen above appears the first time the program is started. Speech Studio provides a wide range of choices that are designed to support work in the clinical environment. This Main Display is made up from a number of these interactive windows. Each window can be removed or selected so as to suit the special needs of clinician and patient. When Speech Studio is subsequently started, the set-up existing at the time the program was closed is used.

Clicking **View** on the menubar in the top left of this screen selects a menu that gives access to all the main windows of the first display. If they are ticked they will be shown. If you click on a tick the corresponding window is removed. To obtain the very simplest display untick all the window names in the menu shown left.

The Parts of the Main Display Screen



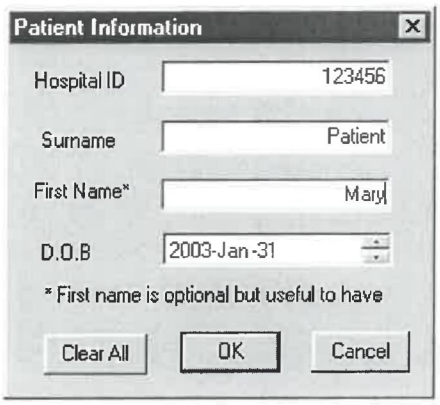
The following pages of this guide are designed to assist the user of Speech Studio to make use of all these facilities – and many more – by starting from the simplest displays and building up to particular applications for individual users. A good screen for beginners to start with can be obtained by clicking **View** then clicking **Toolbar**, **Status Bar** and **View Selection Bar**. On the selection bar check only the **Waveform** box.

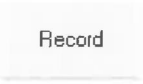


Recording Data

Before recording a new patient set the white dots on the blue and green control knobs on the Processor to the 12 o'clock position. Attach the electrodes to the patient and ask him/her to phonate for a few seconds. If the green LEDs next to each control knob light then the attachment is good.

The next step is to enter patient details and this is done by first clicking the patient icon (right). A new panel is then superimposed on the Main Display Screen to facilitate the entry of file details. Data entries for Hospital ID and Surname are mandatory. Click **OK** when the data is correct. This step is essential as it sets up a unique data folder for the patient consisting of the surname and ID i.e. Patient_123456. A sub-folder of the current date is also created e.g. 20021201. When recorded data is saved it will automatically save into lxdata\Surname_ID\yyyymmdd. Thus each file is stored in a patient and date identified location.



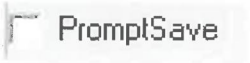
Next click on the Create New Document icon . A new window opens in the central display area. Maximise the window by clicking the square in at the top right of the window. To start a recording press the **spacebar**, click the icon or the  button. There are two recording modes: single screen and continuous. If **SingleScreen** at the bottom of the **Selection Bar** is checked the waveforms will cross the display once then stop. If it is not checked, the patient information box will appear again to ensure you have entered the patient data correctly. If so, click **OK** and a continuous recording starts in which the waveforms cross the display then start again from the left. If the patient data is not correct it may be changed now before clicking **OK**. If **Cancel** is clicked after changing the data, the changes will not be saved and the recording will then start with the original patient information. Either mode of recording can be stopped by pressing the **Enter** key, clicking the icon or the button. For the continuous mode a maximum recording time can be set – see the note on the use of **Set** (page 15).

Zoomed Wave Mode

To check the waveforms before starting a clinical recording open a new document and from the **View** menu click **Zoomed Wave Mode** (right). Start a recording and the display will continuously show only a few cycles of the Sp and Lx waveforms. The control knobs on the Processor can now be adjusted to ensure good waveforms. Stop the recording and close this window without saving the data before doing any clinical recordings.

Saving Data

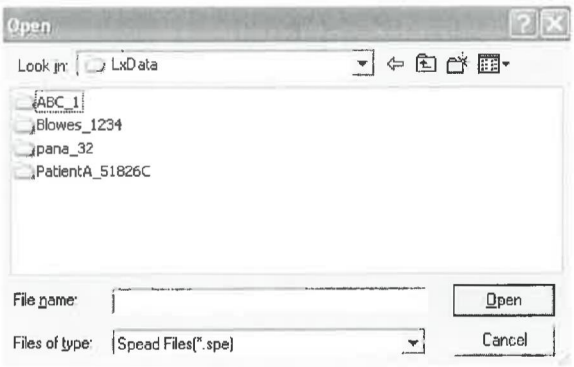
If you wish to keep the data from a clinical recording it is essential to save it. If you do not save the data and start another recording in the same window you will lose the previous data. To save the whole recording click the icon. The Save dialog box (above) appears with the directory automatically selected according

to the patient surname, ID and test date as discussed on page 3. A filename is suggested by the program. If this filename is satisfactory click **Save**, the dialog box will disappear and the data is saved. Alternatively enter your own file name and click **Save**. Clicking **Cancel** means the data is not saved. To save only a section of the data see page 9. If the **Prompt Save** box 

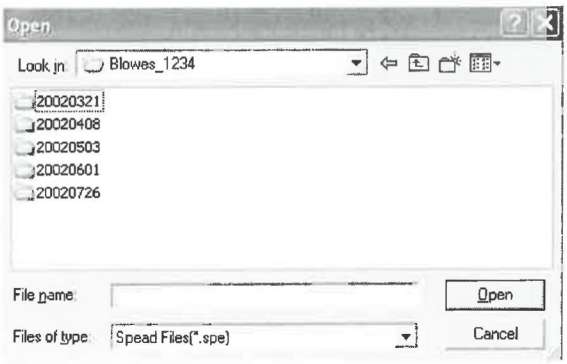
at the bottom of the **Selection Bar** is ticked then after a recording is stopped the Save dialog box appears automatically. This feature is useful if data is being lost due to forgetting to save it. If data has not been saved and the user tries to close the display window or the program itself, again the Save dialog box will appear. If the data does not need saving click **Cancel**.

Opening a Saved File

Going to **File** in the top left hand corner of the main screen and selecting **Open** or clicking the icon gives access to the saved data files via the Open dialog box.

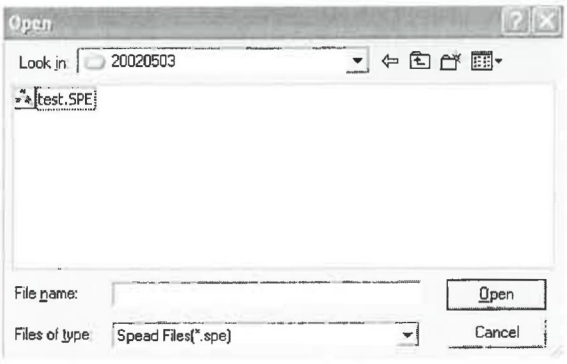


Select the directory where the Speech Studio files are stored (usually c:\lxdata or d:\lxdata) and an alphabetical list of patient folders will appear (top right).



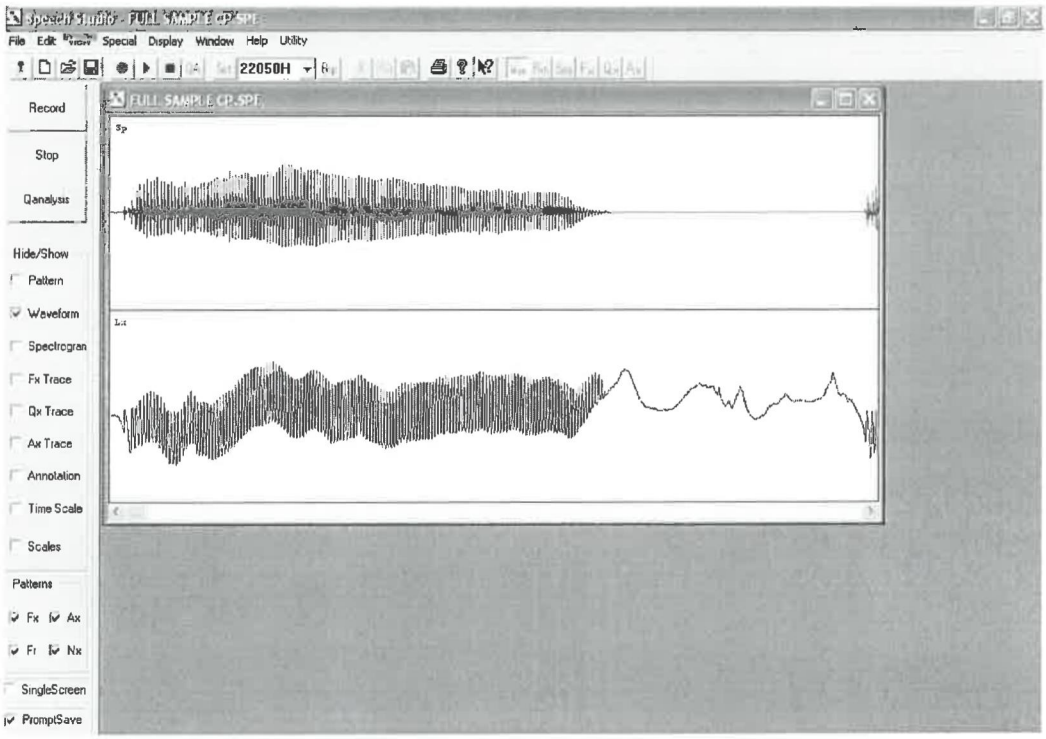
Double clicking on a folder will show subfolders of the dates on which recordings were made (middle right) and double clicking one of the dates will reveal the files saved on that date (bottom right)

Double clicking on a file name will open the file and display the data and in the central display area.



Displaying Recorded Data

Once recorded data has been saved or a previously saved file has been opened the central display area will show a portion of the complete recording. In the example below the data is in a large file and only the first utterance in the complete recording of the waveforms is displayed in this window. The following sections show how to listen to, access and measure all the data.

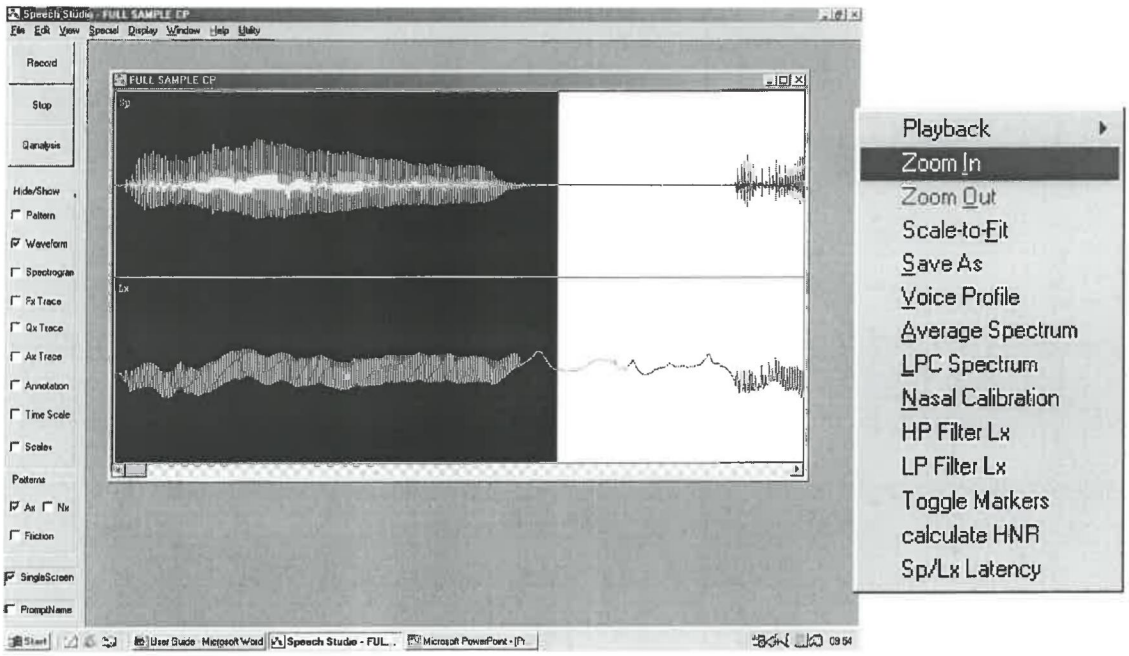


If the data window doesn't fill the central display area, as above, maximise the window by clicking the square in or double clicking anywhere in the blue area surrounding it.

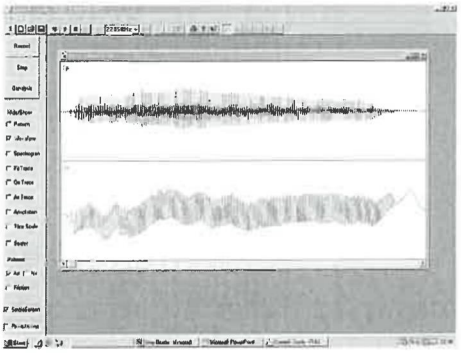
Zooming and Viewing

Zoom In

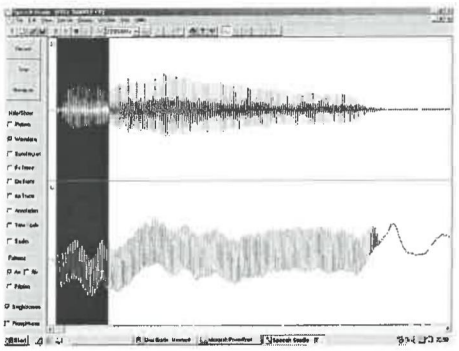
It is possible to examine any aspect of the waveforms in greater detail by blocking in a section and clicking on **Zoom In** on the context menu. Blocking in is done simply by holding the left mouse button down whilst dragging the mouse over the section of interest and then releasing the left mouse button. To access the context menu click the right mouse button while the cursor is anywhere within the blocked-in area. From the example above the first utterance has been selected in this way (below).



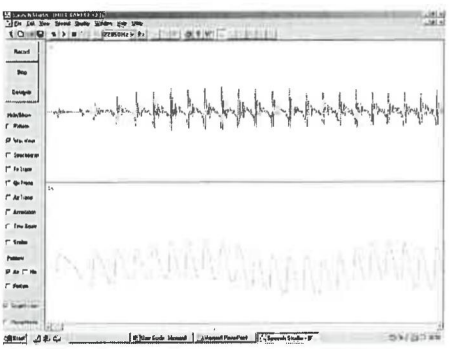
This operation results in the display right.



The next display results from maximising the window by clicking on the square in the top right hand corner of the waveform window.



Using the mouse to block in the onset of the vowel waveform and using the **Zoom In** option in the drop down menu it becomes possible to examine the details of vocal fold initial vibration and acoustic response. The unusual long onset closure is easily seen.



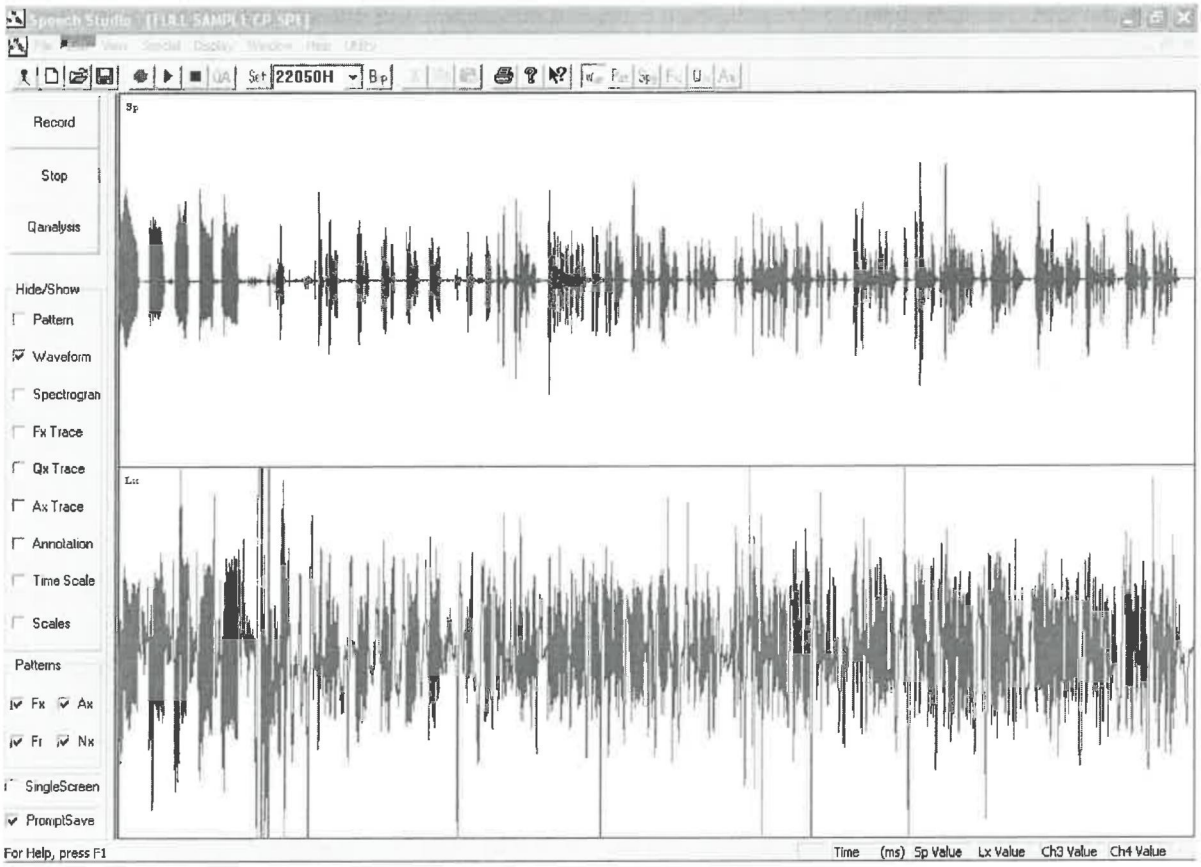
Zoom Out

Once a Zoom In operation has been completed right clicking anywhere in the display window and selecting **Zoom Out** will take the window back to the previous level of zoom.

Using Scale to Fit

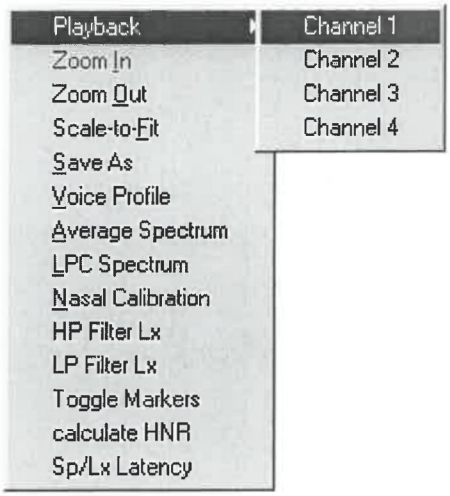
The vowel sounds shown in the preceding displays have both been taken from a complete session in which vowels, citation sentences and a read passage were recorded. Using the Scale to Fit function it is possible to view the whole of the session on a single display.

Set the mouse pointer is set anywhere in the waveform display and click the right mouse button to access the drop-down menu. Left click on **Scale-to-Fit** and the whole of the recording will appear in the central window resulting in the display below.



Listening to the Recording

To listen to the whole of the speech recording press the icon on the toolbar. Playback can be stopped by pressing the icon.

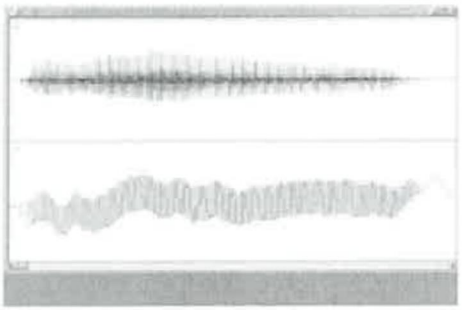


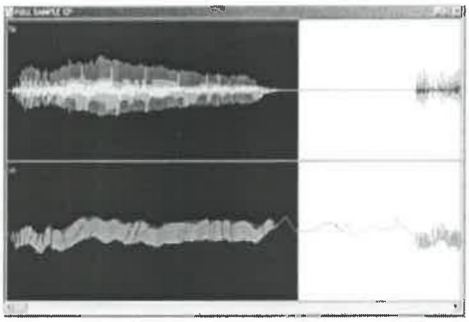
To playback a section of the recording use the Playback option from the context menu as described below. To choose a section either **Zoom In** to the area of interest and right click anywhere in the window or block in the section of interest and right click in the blocked area. When the right button is clicked the menu shown right appears.

If the option **Playback** is chosen, followed by **Channel 1** from the second menu on the right, then the top speech waveform, Sp, is played out. In this example there are only two channels and the choice of **Channel 2** will play out the lower waveform, Lx, and a sound from the vocal fold vibration detected by the laryngograph is heard.

Saving a Section of Data

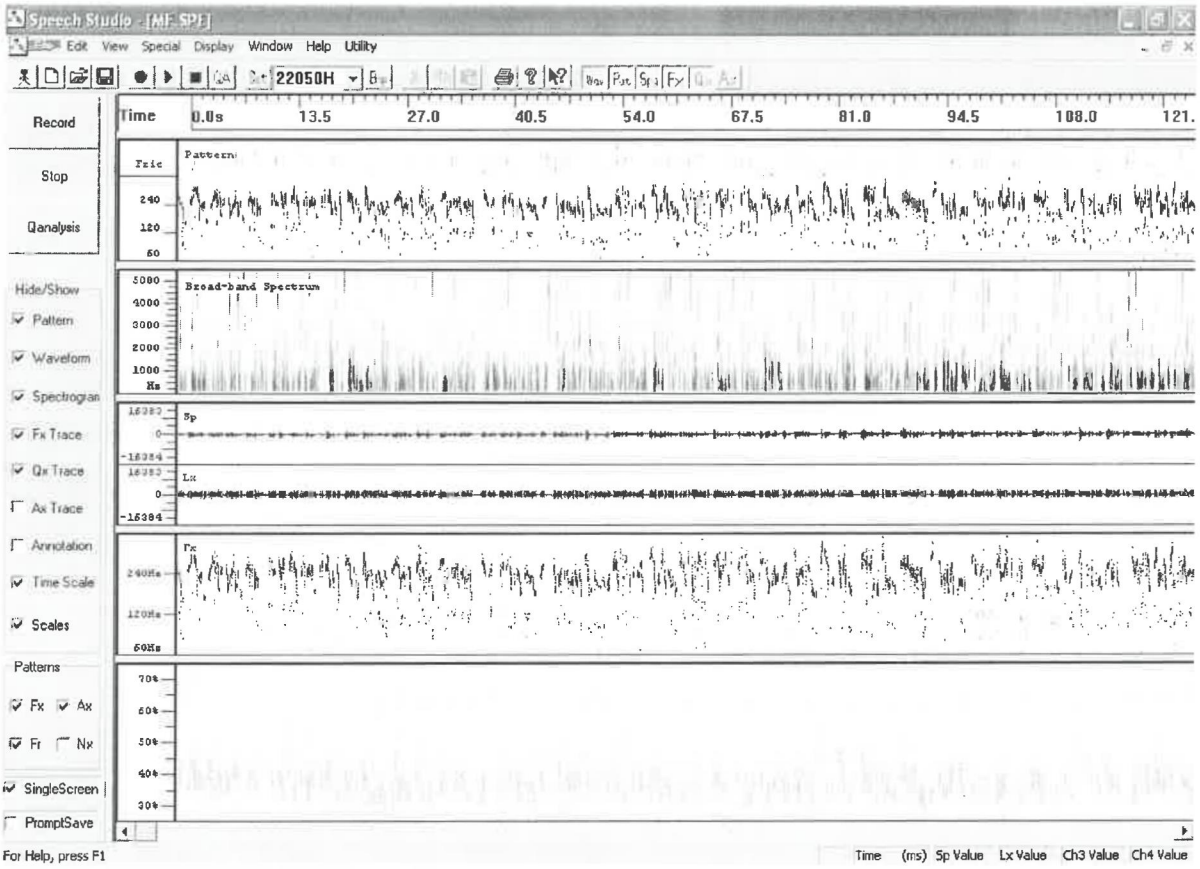
Rather than saving the whole recording by clicking the icon, a smaller section may be saved using the Save As function from the context menu. The section to be saved may be selected either by zooming in so the waveforms are displayed in the display window (above right) or blocking in the required portion from a longer recording (below right). In the first case right click anywhere in the window, select **Save As** and enter information as described above. In the latter case right click within the blocked in area and then choose **Save As**.





Selecting Displays

Displays other than **Waveform** can be viewed on the central display area. On the **Selection Bar** (right), clicking on the respective button will show that parameter if not currently displayed (button not ticked) or hide it if already displayed (button ticked). In this example **Waveform** is ticked so only the Sp and Lx waveforms would be displayed. The display below demonstrates the result of ticking all the boxes (Ax Trace and annotation are not functional).



Pattern displays the voice pattern waveform (see below).

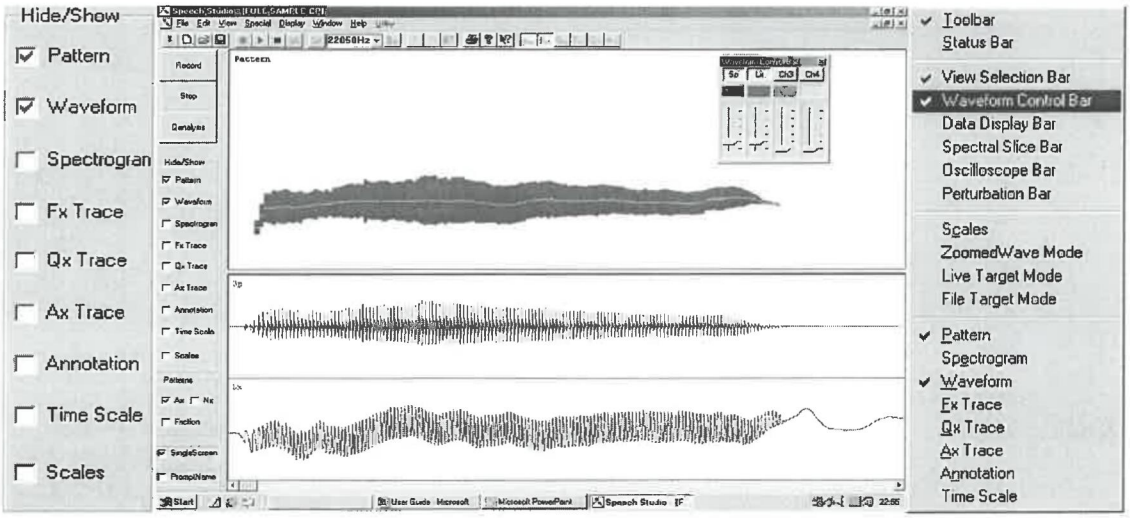
Spectrogram displays a real-time plot of intensity against frequency. Once recording is finished use the context menu to toggle between a broadband and narrowband display and to set the vertical resolution.

Qx provides a continuous display of closed quotient and similarly **Fx** a display of fundamental frequency (this is the same as the Pattern waveform with only the Fx component).

Time Scale puts a scale in seconds across the top of the display window.

Scale puts vertical measurement scales on each parameter window.

Voice Pattern Display and Waveform Control Bar



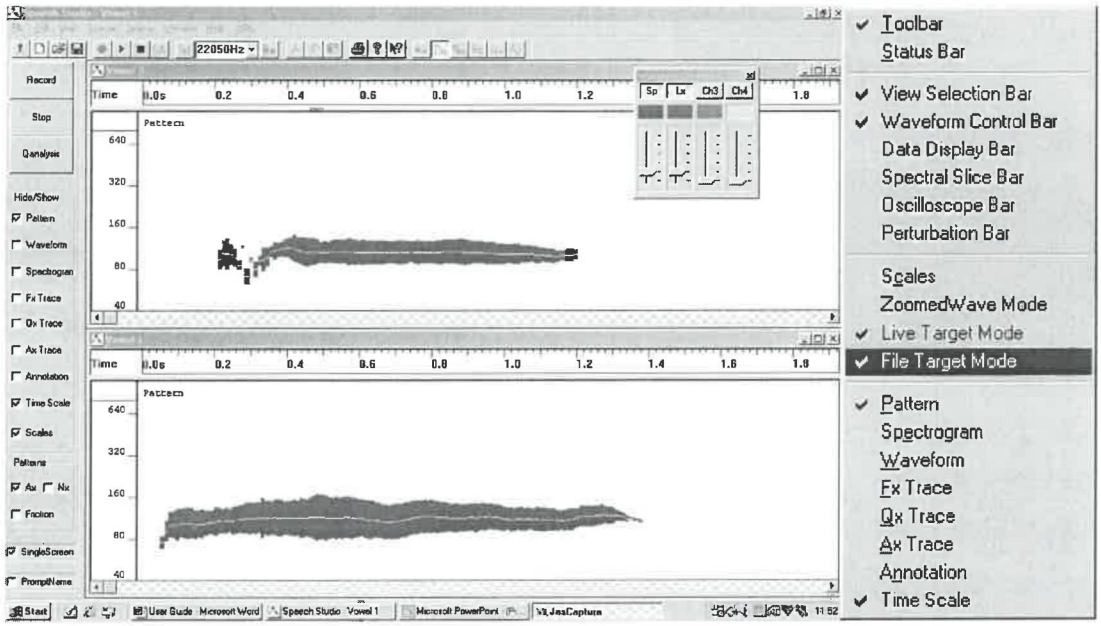
In this example the **Pattern** option has been chosen from the **Selection Bar** and the **Waveform Control Bar** has been clicked from the **View** menu. The sliders on the inset panel now make it possible to change the size of the waveforms. They also control the vertical height of the **Pattern**, this display shows the pitch of the voice from the central thin line and the loudness by its height. **Note that the sliders do not alter the size of the recorded waveforms, only the size of the displays. It is essential that the recorded waveforms are of sufficient amplitude otherwise the software will be unable to analyse them (see Recording Data on page 3).**

Pitch comes from the period by period measurement of vocal fold vibration **Fx** and loudness is measured from the peak to peak height of the **Sp** waveform at the start of each period. In this way the timing control of voice is shown without any smoothing.

There are four possible components to the **Pattern** display – pitch (**Fx**), loudness (**Ax**), Frication (**Fr**) and Nasality (**Nx**). **Fx** is always displayed but the other three components can be shown/hidden as required. Near the bottom of the **Selection Bar** is the **Patterns** control box (left). When a box is ticked that parameter is displayed on the **Pattern** display. In this example only the pitch contour would be displayed. **Nx** is only available with a Nasality Processor.

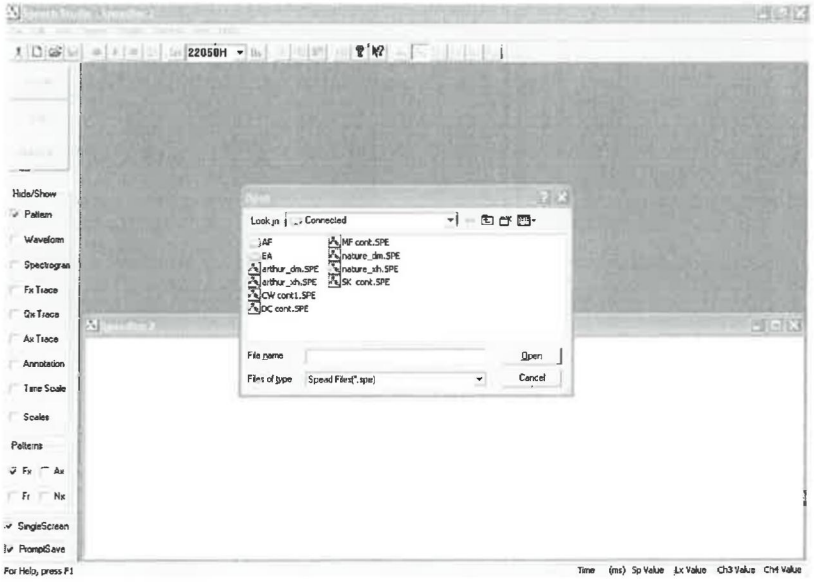
Target Modes

Target mode is a biofeedback mode. There are two ways of using it. Clicking on **View** then **Live Target Mode** opens two windows arranged one above the other (below). The top window is nominally designated the therapist's trace and is initially active. On the **Selection Bar** check **Pattern** only and then record a sample trace for the patient to copy. A recording can be made in this window repeatedly until a satisfactory result is obtained. Now clicking anywhere in the bottom window makes this display active. Select the same setup as for the upper trace whereupon the patient can record his/her attempt to copy the therapist's trace, again until a satisfactory result is obtained.



The two **Patterns** above give a particular example of one of the ways that this type of voice analysis can be of use. By only showing aspects of pitch and loudness the two utterances can be compared, and used interactively, to concentrate attention on two of the most salient aspects of voice control – and voice perception.

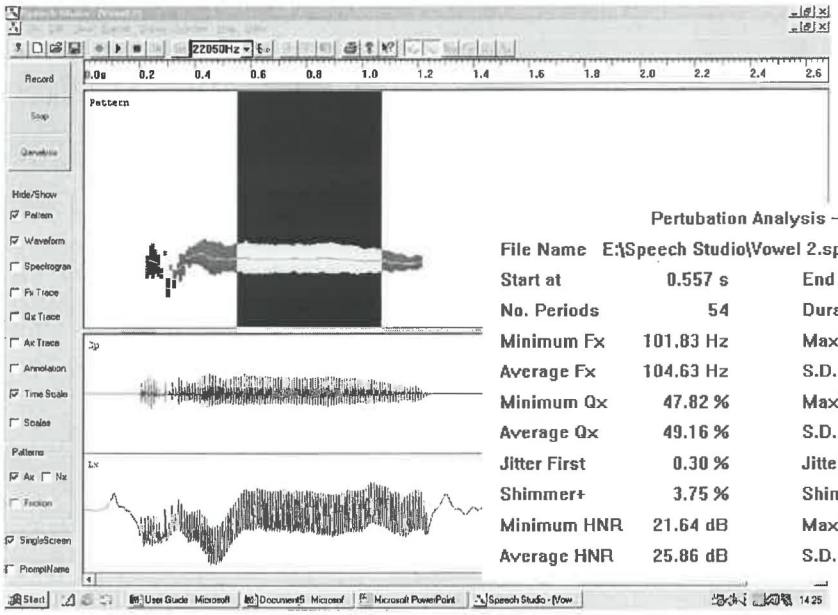
The second **Target Mode** is obtained by clicking on **View** then **File Target Mode** – this results in the screen below.



The patient's window is placed as before but instead of an upper window for the therapist to record a trace, a previously recorded trace can be opened into the top display via the Open dialog box.

Jitter and Shimmer Analysis on Sustained Vowels

To make a measurement on a sustained vowel block in the required section as shown below, right mouse click within the blocked area and then choose **Voice Profile** from the menu.



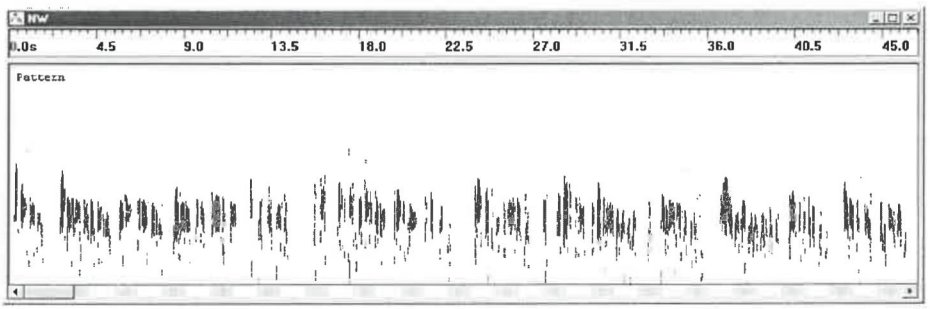
Pertubation Analysis – Holt

File Name	E:\Speech Studio\Vowel 2.spe		
Start at	0.557 s	End at	1.065 s
No. Periods	54	Duration	0.508 s
Minimum Fx	101.83 Hz	Maximum Fx	106.84 Hz
Average Fx	104.63 Hz	S.D. Fx	1.34 Hz
Minimum Qx	47.82 %	Maximum Qx	50.72 %
Average Qx	49.16 %	S.D. Qx	0.07 %
Jitter First	0.30 %	Jitter Second	0.14 %
Shimmer+	3.75 %	Shimmer-	-5.75 %
Minimum HNR	21.64 dB	Maximum HNR	28.82 dB
Average HNR	25.86 dB	S.D. HNR	0.00 dB

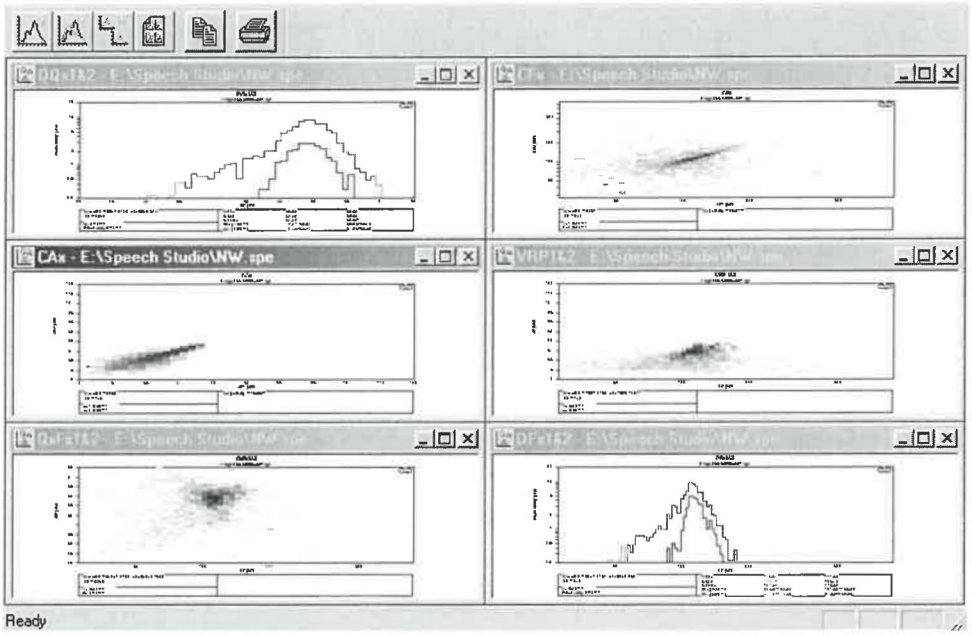
This gives a set of direct measurements on the blocked region. This perturbation analysis can be printed out simply by clicking the icon. Details of how jitter and shimmer measurements are made and what they mean are given in the appendix.

The Use of QA in the Analysis of Connected Speech

The toolbar at the top of the main Display screen has an icon that provides a direct way of producing a set of analyses of the displayed connected speech data. **QA** represents **Quantitative Analysis** and gives access to predefined groups of analysis procedures that are designed for clinical application and which can be tailor-made for the needs of a particular application or of a particular user. Once set it is only necessary to click on the **QA** button to produce the complete group of related analyses. The example immediately below gives a typical **Pattern** display for a read passage.



Choosing **QA** with **Summary** not selected (see The Use of Set below for QA options) gives:



Please see the appendix for further information on QA measurements. A glossary of analyses is also available.

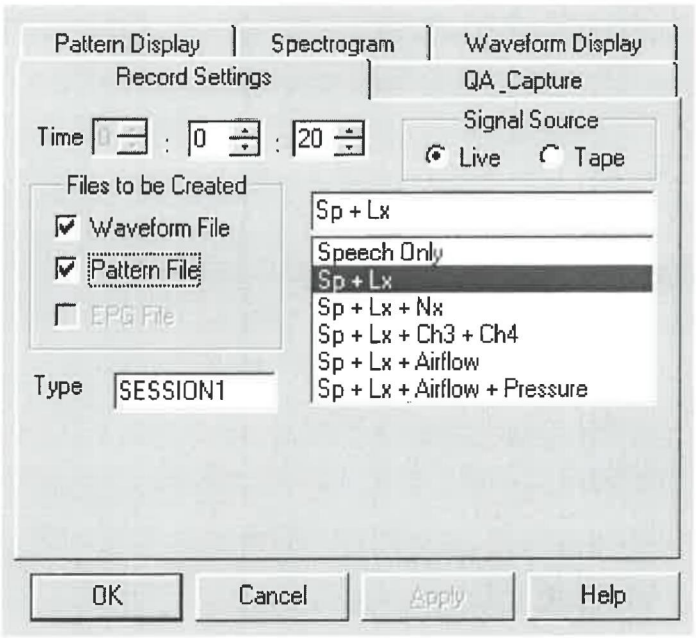
The use of Set

Immediately to the right of the **QA** button is **Set**.

This button is not intended to be in continual use, rather it provides for the initial adjustment of the recording, display and analysis facilities of the Speech Studio system as a whole so that it best serves the needs of patients and users in a particular environment. Clicking on **Set** gives access to five windows.

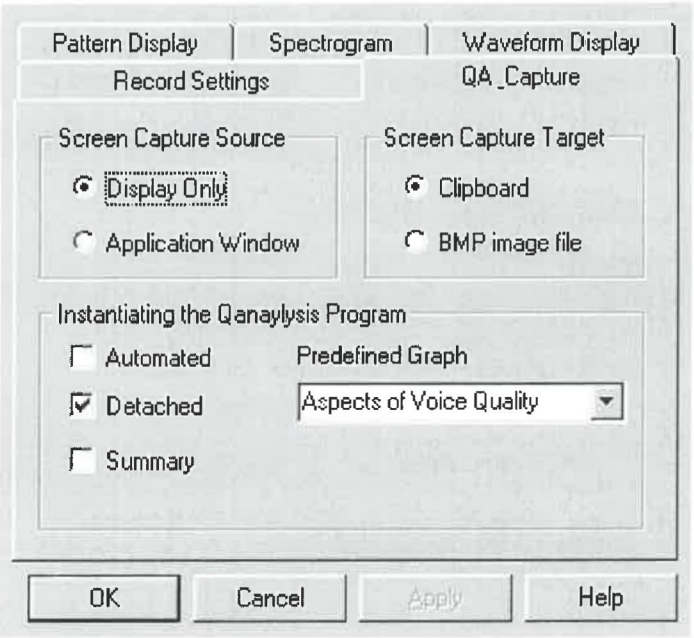
Record Settings

This defines the data channels to be acquired and the maximum duration (h m s) of a continuous acquisition.Two channels, **Sp+Lx,** are set to be acquired - this corresponds to the example data files in the first part of the manual. When using a nasality processor set **Sp+Lx+Nx**. The time limit is set to 20 seconds but in most cases should be set to at least twenty minutes. If recording from a tape set the **Signal Source** to **Tape**.



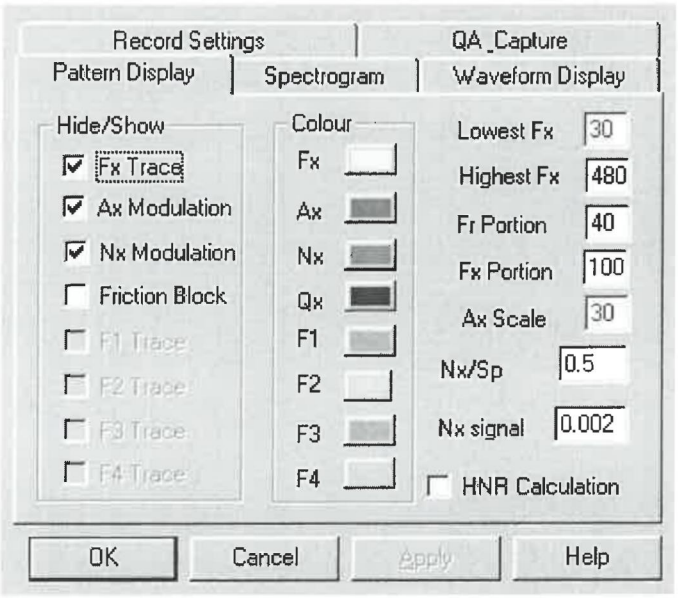
QA Capture

Use this to define the nature of the analyses that are produced when the **QA** tab is selected. To choose a group of analyses use the **Predefined Graph** selection box. of the type of presentation. **Detached** should remain checked. Checking **Summary** puts all the analyses in a complete group on one sheet ready to print as a report. If it is unchecked each graph appears as a separate window.



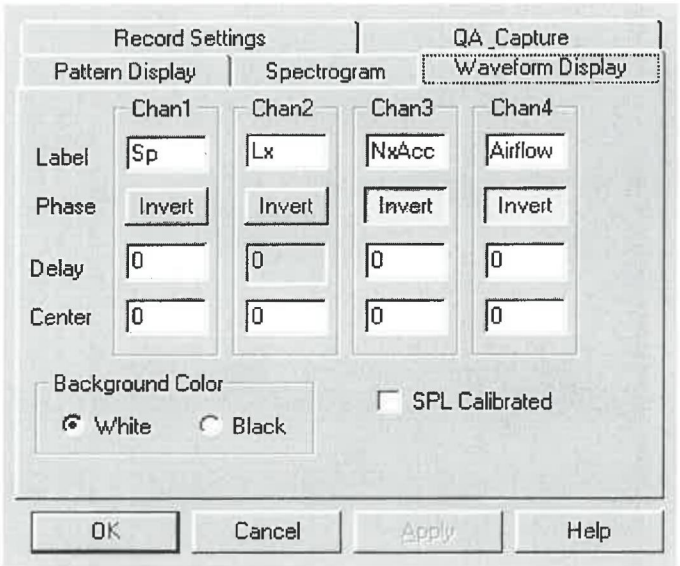
Pattern Display

This window defines data presentations. Once more it is not ordinarily necessary to access this window since the choice once made should be left untouched. The main choices, going from left to right, concern: the data types; the colours associated with the individual analysis traces; and the numerical ranges for each of the components analysed. Tick **HNR** to enable this analysis on the Voice Profile report.



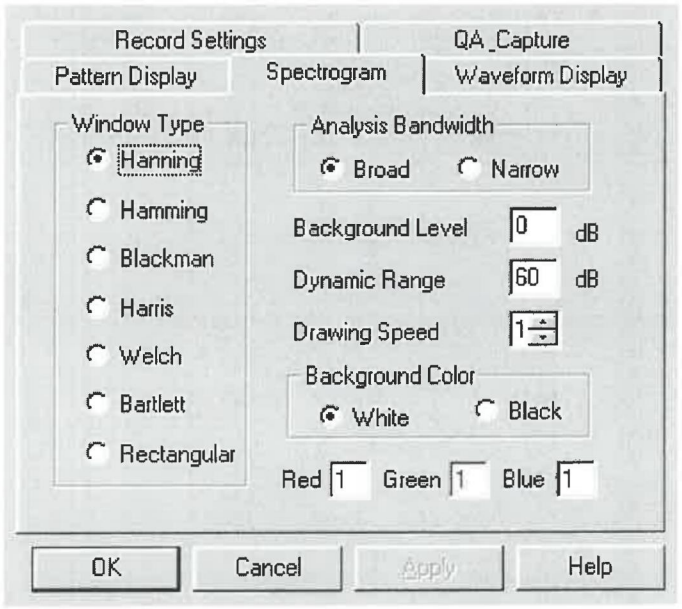
Waveform Display

The fourth window defines the use that is made in data displays of the four channel capability of Speech Studio's acquisition card. Up to four signals can be acquired. **Sp** and **Lx** have been used here. **NxAcc** relates to the nasality display and **Airflow** is for the associated use of mask based sensors. The invert buttons should not be enabled (as for Chan1 and Chan2) and the delay and center values should be 0. Tick **SPL calibrated** if the microphone is to be calibrated with a sound pressure calibrator.



Spectrogram

The fifth window defines the ways in which it possible to set the system for the analysis of spectrograms – real-time plots of intensity against frequency. The two most important parameters that must be set are: for the frequency width of the analysing filter **Broad** *for formant displays* as opposed to **Narrow** *for pitch line displays;* and the two **dB** settings for dynamic range.



Appendix

Laryngograph Speech Studio Analyses

Summary

The aim is to get down to simple basics – and to provide a straightforward set of clinically useable quantitative acoustic analyses that reflect important aspects of what is so obvious to the ear of the listener.

Loudness, pitch and quality are widely used to describe essential aspects of a speaker's voice. However, their basic links with simple parameters of auditory processing are currently little used either in analysis or therapy. In the present work the application of pitch perceptual criteria is described in the provision of an integrated framework for practical clinical assessment and therapy. Sustained sounds are shown to require quite different levels of sampling accuracy from those needed for ordinary connected speech. Radically useful results come from the application of these basic ideas to the representation and analysis of the pathological and the normal voice in case management and audit. Even very abnormal speech samples can be seen to have measurable structures of pitch, loudness and quality – in the midst of apparent disorder.

Examples are discussed of the application of these approaches to pathological voice samples from clinics in several countries. Results from five main analysis types are examined:–

sustained vowel measurement using standard techniques but with 1MHz period time sampling

vocal fold frequency distributions based on voice frequency difference limen related bin sizes and both first and second order (digram) analyses to show the effect of pitch perturbation

crossplots of vocal fold period to period variability which give an overview of intrinsic structure and provide a base for the measurement of irregularity which takes account of normal intonational variation in the speaker's voice

phonetogram and amplitude distribution analyses of connected speech using both simple and digram plots for ordinary connected speech

closed phase ratio distributions again using first and second order analyses to provide measures of this aspect of voice quality.

Two signal inputs have been used, from an acoustic microphone and an electrolaryngograph. These inputs are also basic to a related development in stroboscopy. Although sustained sounds depend on different mechanisms of auditory monitoring and productive control, data from connected speech analysis can be of vital help in guiding the design and clinical use of new stroboscopic equipment.

Using pitch perception to guide voice measurement

For most practical purposes the really important aspects of voice are those that can be heard, and the dominant dimension in hearing voice is pitch. This simple concept leads to the possibility of using some simple quantitative criteria to detect and quantify the differences between "good" and "bad" voices.

Classically, pure tones provide a basic reference for both the definition and perceptual investigation of pitch. Subjective psychophysical data have been stably established over many years. Maximum discriminability is reached between 1 kHz, C6, near the top of the soprano register, and 2kHz with an average best just noticeable difference, jnd, of about 0.7% at 200Hz and 0.4% or 4 Hz in the region of 1kHz [1,2] with individual jnd sensitivities going down to 0.1%. Auditory pitch detection for the frequency ranges of the speaking and singing voice appear to employ mechanisms which operate on the basis of temporal processing [4]. This level of pitch discrimination implies an average ability to detect temporal differences between successive periods of about 4 μs, and for some individuals, 1μs. This temporal signal processing ability for pitch perception is paralleled in auditory lateralisation where interaural time differences of about 2μs to 10μs are detectable.

For steady complex tones and vowels in the fundamental frequency range of conversational speech, the pitch discrimination jnds are even smaller than those obtained with pure tones. Wier [3] and Moore [4] for example, within the range 200 to 600 Hz, reported jnd values from about 0.15% to 0.3%. For vowel-like sounds with simple changing fundamental frequency contours, however, the ability to perceive differences in fundamental frequency is drastically reduced and the jnd may be 8% at about 100 Hz [5]. This increase in, and magnitude of, jnd has also been found for whole word utterances with simple intonation contours, the jnd here never being less than 6%. When more complex contours are used, the differences needed to achieve reliable detection may be as great as 20% [6]. The subjective results for these sound types are not as well established as for sustained sounds and there is a dependence on the duration of the tone. There is, however, a good working

consensus between a large number of reported observations [t'Hart et al]. These established observations give clear implications in respect of the accuracy criteria which should be aimed at for the analysis of the separate categories of sustained sounds and connected speech.

A basic set of tools for accurate voice pitch measurement

Tool number 1

Most methods of voice pitch analysis depend on the use of the acoustic signal of speech sampled at low rates which do not correspond to the requirements imposed by the pitch dL performance of the ear. The essential need which has to be met is illustrated by the graph below.

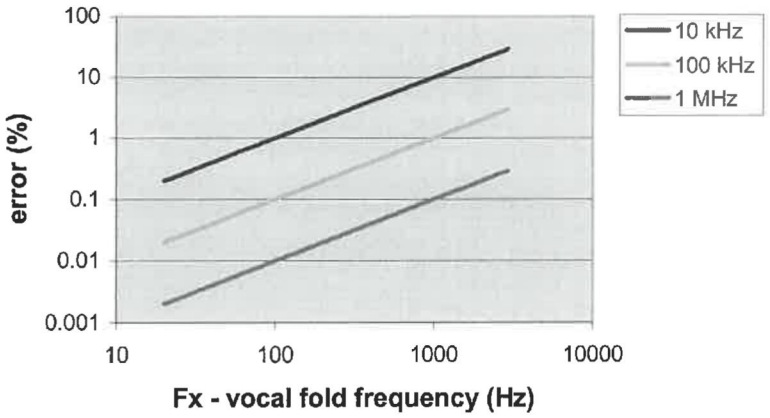


Figure 1 **Voice pitch measurement error and sampling rate**

At a voice frequency of 1000Hz in the soprano range the human ear can detect changes of around 0.1%. In order to do as well as this it is necessary to use a sampling frequency of 1MHz. This is what is done for the following measurements.

Tool number 2

A second basic problem associated with conventional approaches to voice analysis comes from the inadequacy of pitch extraction algorithms based on the acoustic signal. A more reliable technique is to use the electrolaryngographic [egg] output from the speaker's voice activity.

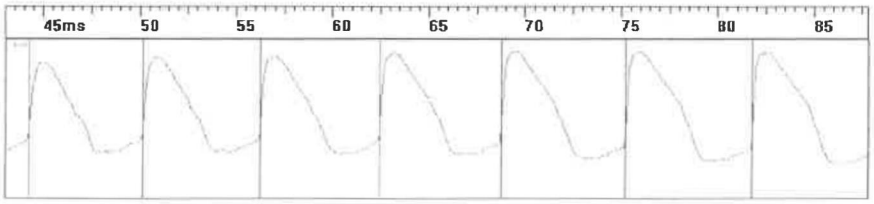


Figure 2 Electrolaryngograph voice signal, Lx, [EGG] sample with period markers

The figure shows a standard Lx/EGG waveform with period markers superimposed. These markers give the basis for an accurate determination of each individual pitch period, Tx, and they can be sampled at 1MHz to give a basis for measurement which, although very highly detailed for many purposes, is linked to the best that the ear can do and provides for considerable flexibility in computational manipulation.

It is, of course, quite easy to deal with sustained sounds but the method must also be reasonably robust when applied to the rapidly changing waveform of running speech. A typical output for a practical system is shown below.

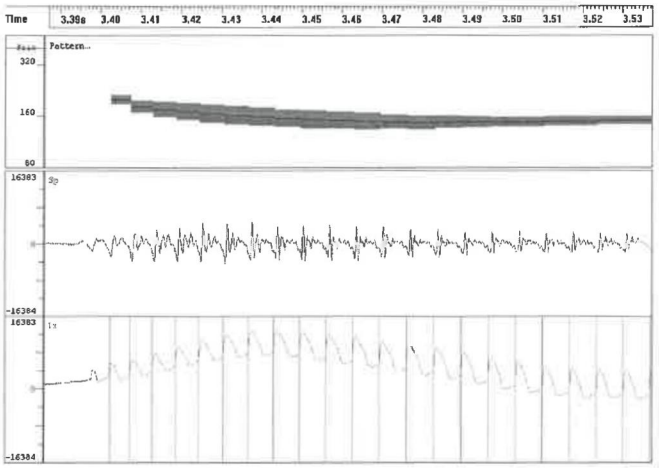


Figure 3 Voice period, Tx, markers for changing pitch and larynx height

The figure shows the same process of marker generation as above, in Figure 2, but now the speech acoustic waveform, Sp, is included and the excerpt is from a sample of fluent speech. At the top of the figure the period by period "instantaneous" frequency, Fx, is shown, with an update for every vocal fold period. The vertical thickness corresponds to loudness being determined by the peak acoustic pressure for each period – a technique which is very useful for interactive clinical work in voice training.

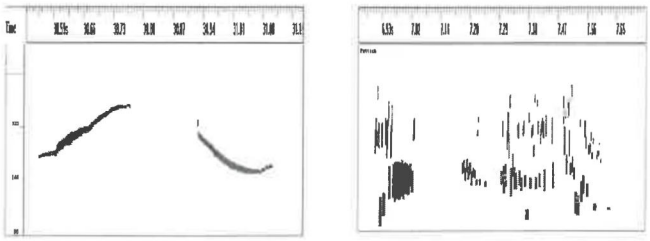
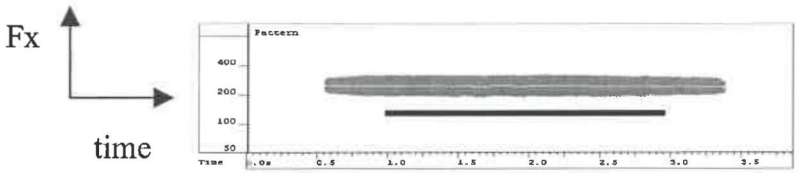


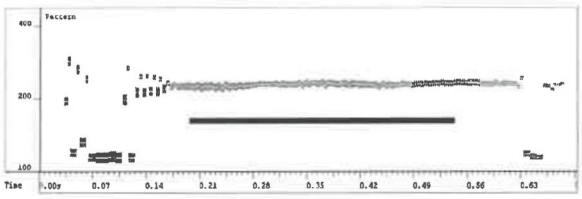
Figure 4 "Pitch and Loudness Patterns" for a normal, left, and an abnormal voice, right.

The use of period by period sampling gives a very clear view of the difficulties that may be encountered by a speaker with a voice disability and shows the remarkable precision of normal voice pitch control. This type of precision data analysis is also basic to the provision of measurements both for sustained sounds and for the analysis of running speech.

The analysis of sustained sounds



Speaker A



Speaker B

Figure 5 Patterns for sustained vowel sounds produced by normal, A, and pathological, B, voices

The current clinical techniques for the quantification of voice abnormality depend to an appreciable extent on the use of sustained sounds and the standard protocol uses the steady state in the centre of the sample. Period by period analysis gives a clear indication of the onset and offset transients which this approach misses. The pathological speaker in general has difficulty in producing smooth voice onset and offset. This is clearly seen initially, where diplophonic breaks in the voice precede more steady production, and in the voice breaks at the end.

speaker	jitter %	dL %
A normal	0· 3	0· 15 to 0· 3
B pathological	0· 8	

Table 1 productive jitter and perceptual dL comparisons

As must be expected, perception leads production but it is striking and commonly observed, that the pathological voice does not have a jitter commensurate with the disability. This relatively small difference results from the choice of the centre interval of a sound sustained at a comfortable pitch – and the speaker's auditory monitoring ability for sustained sounds.

Connected speech and sustained vowels

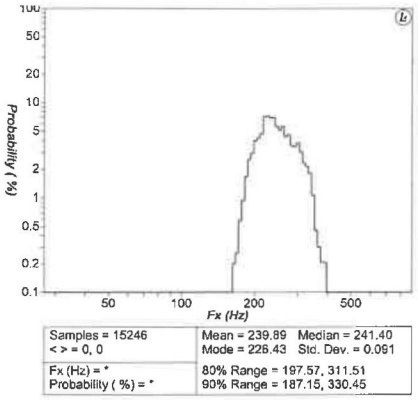
For the majority of the population, speech communication is at the heart of our daily lives. Clinical voice measurement, however, is mostly directed towards the appraisal of the ability to produce a sustained vowel.

Since there are quite substantial perceptual differences between our ability to hear pitch regularity in sustained vowel sounds as opposed to fluent speech, it would be of interest to make at least an initial appraisal of the ways in which perception and production may interact in the voice pitch structures of the two types of phonatory activity. There may additionally be an advantage in comparing pitch regularity inspired analyses based on the two types of spoken material simply with a view to contributing to filling the gap between clinical indices of severity of dysphonia based on vowel measurement and those using a perceptual evaluation of continuous speech. Most important of all, however, is both to make

use of pitch criteria and to take account of the nature of pitched sounds. Regular repetition of an acoustic event and perceived pitch go hand in hand.

Analyses of ordinary running speech

DFx1

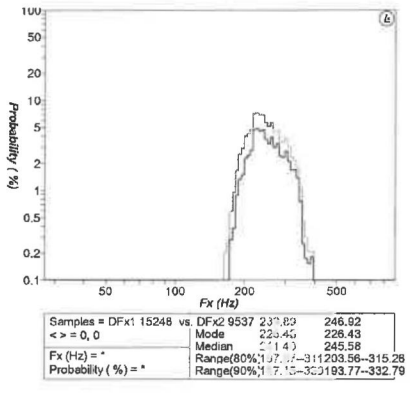


Speaker A

Figure 6 Vocal fold frequency,Fx, distribution for a 2m sample of normal connected speech

This figure shows a standard result for the analysis of the range of vocal fold frequencies contained in an ordinary sample of running speech. The analysis is not, however, standard because it is based on the period by period measurement of vocal fold frequencies with no smoothing. This leads to the possibility of showing and measuring the extent to which the voice contains well defined pitch components, because individual periods can be compared with each other.

DFx1 & DFx2

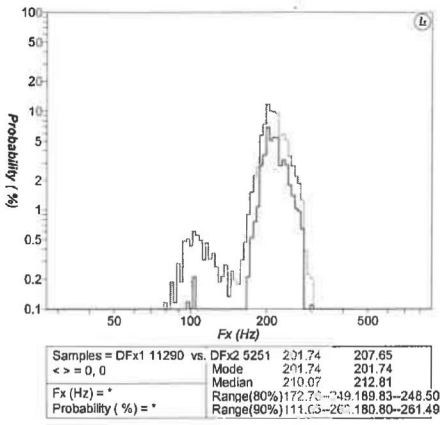


Speaker A

Figure 7 Fx distributions for a whole normal sample (red) AND for its "pitched" components

The use of accurate period by period information makes it easy to plot the occasions when two successive pitch periods have essentially the same value. For the normal voice this happens very often indeed. The pathological voice, however, is very easily identified by the ear as having period to period irregularity. This important feature is shown in the inner of the two distributions above. The two distributions together give an immediate insight into important aspects of voice quality – pitch height and range, modal structure, and regularity. The technique is especially useful for pathological voice analysis.

DFx1 & DFx2



Speaker B

Figure 8 Fx distributions for a whole abnormal sample (red) AND for its "pitched" components

The two distributions here are very dissimilar. DFx1, the outer plot, shows the distribution of Fx values for every vocal fold period in the whole 2m. sample. DFx2 shows only those Fx values for which two successive periods have been essentially the same. Two modes of vocal fold vibration are shown. The main at about 200 Hz is well defined. At about an octave below, the lower mode is more diffuse and is evidently associated with considerable period to period irregularity since the values of DFx1 and DFx2 are so different. Different pathologies give rise to different types of modal structural differences but for most cases the presence of voice pathology will be associated with marked discrepancies in magnitude and shape between these two forms of representation.

Jitter and irregularity in connected speech

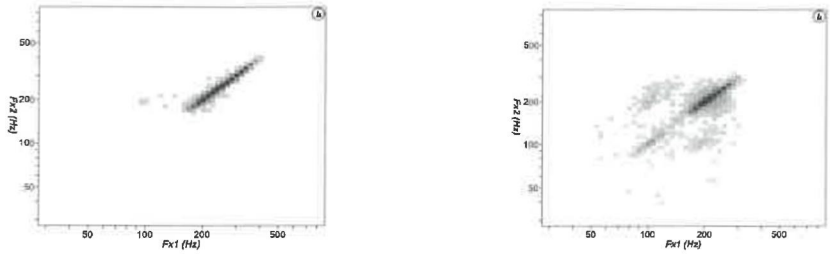


Figure 9 Vocal fold period crossplots, CFx
speaker A on the left, B on the right

Jitter and intonation

The procedure basic to the ordinary application of the jitter criterion is applied only to sustained sounds and requires that the voiced sound being measured is held at as constant a pitch as possible by the speaker. The essential concept, however, is directed at obtaining a quantitative assessment of pitch variability. The idea is just as applicable to ordinary connected speech so as to get an appraisal of the irregularity which may be inherent to the social use of a pathological voice.

An obvious first approach to the measurement of pitch irregularity in a sample of running speech is to determine the standard deviation of the spread of cycle to cycle differences in regard to periods or frequencies. A difficulty with this approach is that it will necessarily include ordinary intonational variations as part of the estimate of irregularity. The problem is perhaps best illustrated with reference to actual data. When vocal fold vibration is essentially regularly periodic the use of a period by period crossplot, as in Figure 9 A, gives a clearly defined diagonal line – since successive periods have almost the same values, apart from the variations arising from the intonational frequency related changes of connected speech. For the pathological voice, however, the shape of the crossplot is not so simply defined because successive vocal fold periods are very often markedly different and are not totally under the speaker's cognitive control. This method of plotting the range of variability in period to period coherence is effectively similar to the application of the jitter criterion, used for sustained sounds, to the whole of a connected speech sample.
The interpretation of jitter in running speech, however, is not at all the same as that for sustained sounds. First, the pitch dLs are quite different in the two cases. The bin sizes needed for the adequate representation of significant changes in the present data involves

6% steps. The 0.1% resolution required for the analysis of sustained sounds is not appropriate. Second, the presence of intonational changes makes it necessary to ignore variations which are part of the normal patterning of vocal fold frequency change in running speech. Figure 9 A shows that there is indeed a centre continuous core of variation for the whole of the vocal fold frequency range and this is found for all normal speakers.

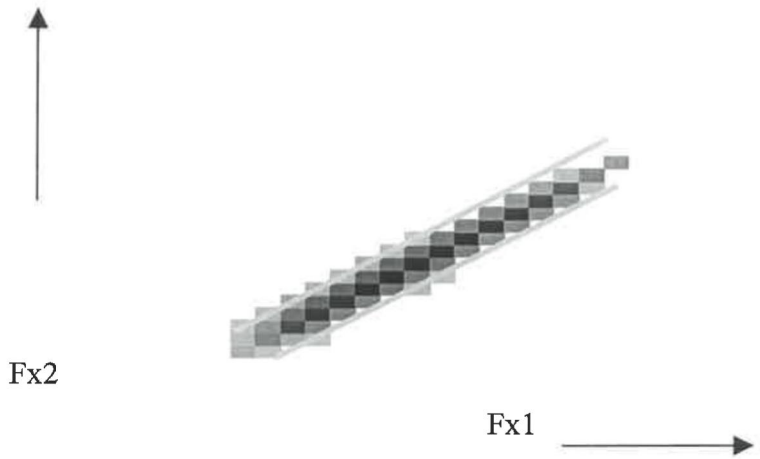


Figure 10 How the core of the normal Crossplot, CFx, is shaped by normal intonation

If the pitch difference limen value of 6% is applied to this data then it becomes possible to apply a theoretically founded criterion which makes it practically feasible to separate the variability arising from intonation from that due to other causes. It is then only necessary to determine all the pitch deviations which are more than 6% away from the centre line in the graph showing Fx1 against Fx2 – where Fx1 is the frequency value of the first vocal fold cycle in any pair of cycles in the whole utterance and Fx2 is the frequency value of the immediately following cycle of the pair. Fx is used to denote the frequency value of a single vocal fold cycle, the period of this cycle being measured from point of closure to point of closure.

Normal and pathological voice examples

The comparison of Figure 9A with 9B shows how the relatively small jitter differences for the sustained sounds produced by these speakers in Table 1 is related to quite marked structural changes in their samples of connected speech. In these particular instances, irregularity is 3· 2% for the normal speaker and 14· 7% for pathological speaker B. Both values were measured in the way described above as a percentage of the number of vocal

fold periods, outside the centre core of intonation-dependent pitch change, relative to the total number of vocal fold periods in the whole spoken sample.

Loudness and Quality

Connected speech phonetogram

The standard phonetogram was designed to provide an overview of the dynamic range of a singer's voice and was based on the separate production of sustained sounds. The same principle can be applied to the analysis of the speaking voice to give a "Dynamic Phonetogram" which is derived from the amplitude–frequency analysis of a complete sample of connected speech.

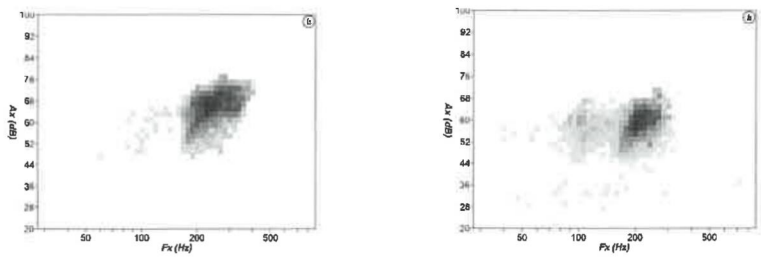


Figure 11 Dynamic Phonetograms derived from samples of connected speech: normal speaker A on the left; abnormal voice speaker on the right

In both Figure 11A and B, the first and second order distributions have been superimposed. This has little effect on the presentation for speaker A; it does have a profound influence on the form of the data presentation for speaker B since the presence of a bimodal peak in loudness is very evident in the first order distribution but not in the "pitch" related second order plot.

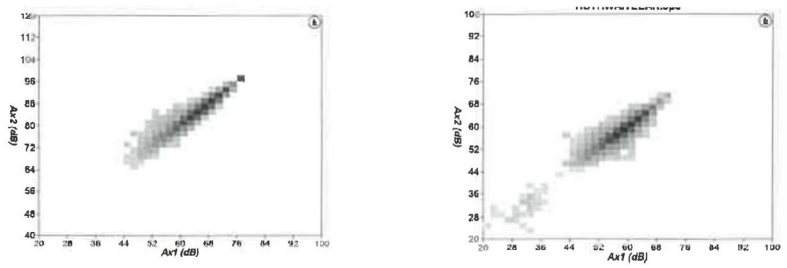


Figure 12 Period by period amplitude crossplots – CAx

A factor contributing to our perception of hoarseness comes from the irregularity of successive amplitude peaks in the cycle to cycle excitation of the vocal tract. This is especially evident in connected speech and speaker A on the left has a smaller spread in these analyses than B. Using a similar measure of irregularity to that employed for CFx gives values respectively of 3· 3% and 6· 5%.

Voice quality, "closed" phase and pitch

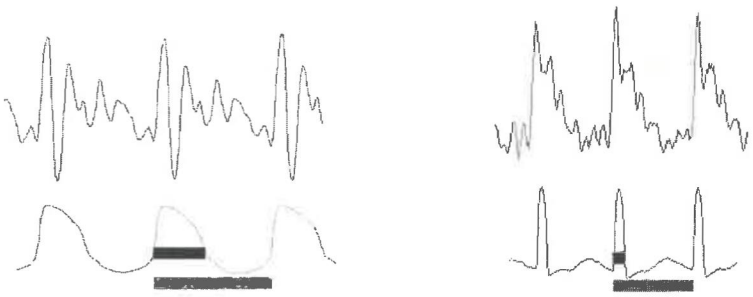
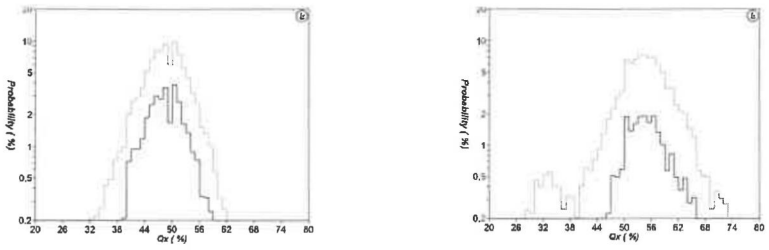


Figure 13 Using Lx transconductance to estimate "closed phase" ratio

A direct method for the estimation of the closed phase of each glottal cycle can be based on the use of the Lx/egg waveform. The black bars in Figure 13 show the closed phase at a point 70% down from the peak of the waveform and this is simply taken as a ratio with the total period, Tx, to give this parameter. In what follows the symbol Qx is used to denote this particular closed phase estimate. As before, Qx can be measured for each individual period and it can also be linked to the pitch regularity which occurs when two successive periods have essentially the same value – to get the digram or second order evaluation of closed phase in running speech.



A B

Figure 14 DQx 1&2 – distributions of first and second order "closed phase" as a function of vocal fold frequency, Fx

Voice quality is a complex attribute of voice but one important aspect comes from the regularity and duration of the closed phase from vocal fold cycle to cycle. First and second order plots can often give important information in regard to the physical nature of a pathological voice and here it is evident that speaker B has poor closed phase coherence.

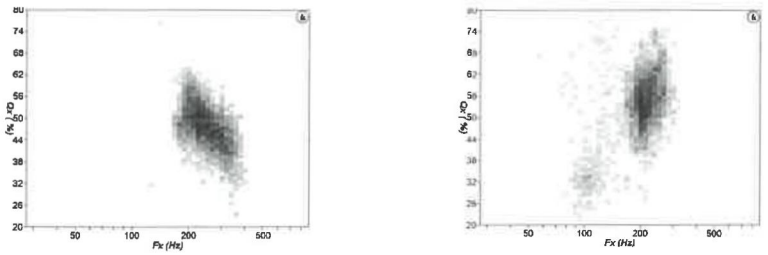


Figure 15 "Closed phase" ratio Qx as a function of vocal fold frequency

In normal speech the variation of the closed phase at the level of discourse is a rule governed phenomenon whose rules differ from one person to another. Speaker A is a young woman and shows an effect of increasing breathiness with larynx frequency (work by Evelyn Abberton). The pathological voice, B, is substantially deviant and gives a range of Qx [the closed phase measure based on trans-glottal conductance] which is never found in the normal voice and shows the irregularity as a function of frequency which can be clearly heard in her speaking voice.

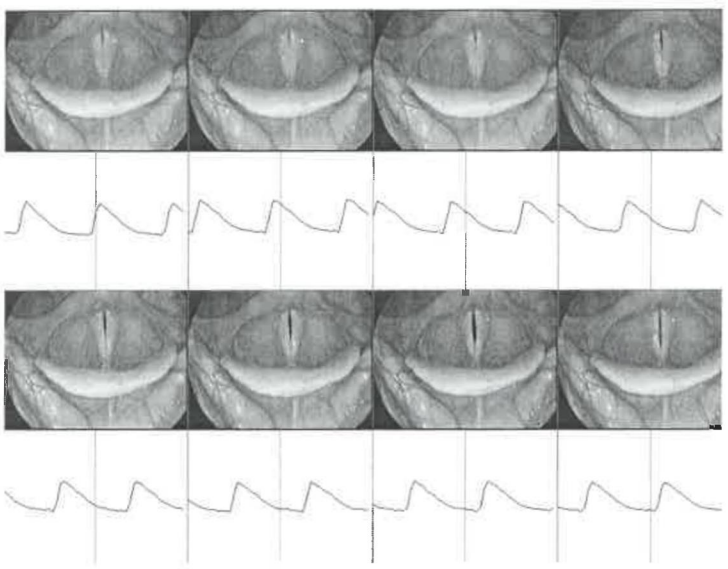


Figure 16 A brief illustration of the use of precision timing in the simultaneous stroboscopic acquisition of vocal fold images and Lx/EGG waveform information.

Figure 16 gives a simple example of the combination of the Lx/egg information that is used for the analysis of sustained and connected speech with stroboscopic images. The use of high precision timing to control the stroboscope is especially valuable in image acquisition. The immediate link to the Lx waveform is an aid to diagnosis. The system also allows the possibility of linking connected speech analyses to stroboscopic interpretation so that modes of abnormal vocal fold vibration can be better defined and investigated.

Acknowledgement

The analyses used have all been derived via the use of Laryngograph Ltd. Speech Studio and LxStrobe hardware and software.

References with particular relevance to pitch perception

[1] Nelson, D.A., Stanton, M. E., and Freyman R. L., 1983, "A general equation describing frequency discrimination as function of frequency and sensation level" *J.A.S.A. 73, 2117-2123*

[2] Vance, T.F.,1914, see page 340 in E.G. Boring, *"Sensation and Perception in the History of Experimental Psychology"*, Appleton-Century-Crofts Inc. New York 1942

[3] Wier, C.C., Jesteadt, W., Green, D.M., 1977, "Frequency discrimination as a function of frequency and sensation level", *J.A.S.A. 61, 178-184*

[4] Moore, B.C.J. Glasberg, B.R., Peters, R.W., 1985, "Relative dominance of individual partials in determining the pitch of complex tones" *J.A.S.A. 75, 550-561*

[5] Klatt, D. H. 1973, "Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception", *J.A.S.A. 53, 8-16*

[6] t'Haart, J.,1981, "Differential sensitivity to pitch distance, particularly in speech" *J.A.S.A. 65, 811-821*

Aronson, A. E. (1985) Clinical Voice Disorders. 2nd edition Thieme Inc. NY

Anastaplo, S. & Karnell, M.P. (1988) Synchronized videostroboscopic and electroglottographic examination of glottal opening. J.A.S.A 83, 1883-1890

References with particular relevance to voice evaluation

Baken, R. J., (1987) Clinical Measurement of Speech and Voice. Little, Brown & Company, Mass USA.

Cowie, R., Douglas-Cowie, E. and Rahilly, J. (1988) The intonation of adults with postlingually acquired deafness: anomalous frequency and distribution of elements. In B.

Ainsworth and J. Holmes (Eds) Proceedings SPEECH '88, 7th FASE Symposium, Edinburgh, **2**, 481-487.

Dejonckere, Ph., Remacle, M., Fresnel-Elbaz, E., Woisard, V., Crevier-Buchman, L., Millet, B. (1996) Differentiated perceptual evaluation of pathological voice quality: reliability and correlations with acoustic measurements. Rev. Laryngol. (Bordeaux) 117, 219-224

Fabre,P., (1957) Un procédé électrique percutané d'inscription de l'accolement glottique au cours de la phonation. Bull. Nat. Méd., 141, 66-99

Fourcin, A, & Norgate, M.(1965) Measurement of trans-glottal impedance. Progress Report, Phonetics laboratory University College London, pp 34–40

Fourcin, A., Abberton, E. (1972) First applications of a new laryngograph Volta Review, **69**, 507-518 {reprinted from Med. & Biol. Illustration 21, (1971), 172-182}.

Fourcin, A.J., (1974) Laryngographic examination of vocal fold vibration. in B.Wyke (ed.) Ventilatory and Phonatory Control Systems. OUP 1974, pp. 315-333

Fourcin, A.J.,(1981) Laryngographic Assessment of Phonatory Function ASHA Reports #11 pp116–127

Gilbert, H.R., Potter, C.R., & Hoodin, R (1984), Laryngograph as a measure of vocal fold contact area. J. Speech Hear Res., 27, 178-182

Hammarberg, B. & Gauffin, J., (1995) Perceptual and acoustic characteristics of quality differences in pathological voices as related to physiological aspects,. in O. Fujimura & M.Hirano (eds.) Vocal Fold Physiology pp. 283-303

Hirano, M., (1981) Clinical Examination of Voice. New York, Springer-Verlag

Jones, C. (1967) 'Deaf Voice' - a description derived from a survey of the literature. Volta Review, **69**, 507-8.

Kent, R.D., (1996) Hearing and Believing. Am. J. Sp-Lang Path., 5, 7-23

Laukkanen, A.-M., & Vilkman, E. Tremor in the Light of Sound Production with Excised Human Larynges in P.H. Dejonckere, O.Hirano, J. Sunberg, (eds.) Vibrato 1995, Singular Publishing, pp. 93-110

Laver, J. (1991), The Gift of Speech Edinburgh Un. Press

Lecluse, F.L.E., (1977) Elektroglottografie. Drukerijelinkwijk B.V.

Maddieson, I., (1984) Patterns of Sounds, CUP Cambridge

Scherer, R.C., Druker, D.G. & Titze, I.R. Electroglottography and Direct Measurement of Vocal Fold Contact Area. in O. Fujimura (ed.) vol.2, Vocal Fold Physiology, Raven Press, pp. 279-291

Moore, G. P., (1971) Organic Voice Disorders. Englewood Cliffs, N.J., Prentice-Hall

1 Foundry Mews, London, NW1 2PE
Tel: +44 (0)207 387 7793 Fax: +44 (0)207 383 2039

lx@laryngograph.com www.laryngograph.com